Data Custodian Skill Assessment Survey

Looking to build deeper expertise as a Data Custodian?

As part of the Data Stewardship Service, we are offering this *optional* skill assessment to help you get an understanding of the skills that will help succeed in your role as a Data Custodian as well as gauge your individual skill level. After completing this assessment you'll receive your results. For the areas in which you'd like to advance your skills, the Data Stewardship Service will provide you with learning path recommendations.

Skill Assessment Overview:
As you move through this assessment, you will first see a set of core data terms. Our team wants to gauge your familiarity with these terms since they are essential to your role. After completing the data terms section, you'll go through **six t skill topic areas** and rate your ability to complete the Data Custodian task areas on a scale of **1- 5**, with '1' being the least knowledgeable, to '5' being the most knowledgeable.

Skill Assessment Topic Areas:

1. Data Terms
2. Enhance and Create Data Flows
3. Metadata Management
4. Reference Data Management
5. Data Integration and Interoperability
6. Golden Record Data Management
7. Data Stewardship

*Note: All skills in this survey have been based on the Federal Data Skills Catalog

Your answers to this skill assessment survey will be confidential only to you and the Service core team. Our core team will use your results to make customized training recommendations for you so that you can grow your skills as a Data Custodian. If you have any questions regarding the survey, please reach out to the CoP Core team at ***datastewardshipservicedevelopment@nasa.onmicrosoft.com.***

may send comments on our time estimate above to briana.hila@nasa.gov. Send only comments relating to our time estimate to this address.

Section 1:

Data Terms

These questions are designed to ensure that as an agency we use the same definition of data terms so that we can work more seamlessly as a data community. If the data term and definition matches with your definition/understanding of the term, select "**Yes**." If it doesn't, select "**No**."

1. **Data Management:** Consists of the practices, architectural techniques, and tools for achieving consistent access to and delivery of data across the spectrum of data subject areas and data structure types in the enterprise, to meet the data consumption requirements of all applications and business processes. Is this the definition you use to define data management in your day-to-day work?
2. **Data Governance:** The specification of decision rights and an accountability framework to ensure the appropriate behavior in the valuation, creation, consumption and control of data and analytics. Is this the definition you use to define data governance in your day-to-day work?
3. **Data Architecture:** Describes how data is managed from collection through to transformation, distribution, and consumption. It sets the blueprint for data and the way it flows through data storage systems. It is foundational to data processing operations and artificial intelligence (AI) applications. Is this the definition you use to define data architecture in your day-to-day work?
4. **Data Integration:** Comprises the practices, architectural techniques and tools for achieving the consistent access and delivery of data across the spectrum of data subject areas and data structure types in the enterprise to meet the data consumption requirements of all applications and business processes. Is this the definition you use to define data integration in your day-to-day work?
5. **Data Quality:** Data that is accurate, complete, consistent, reliable, up-to-date, and fit for a purpose. Is this the definition you use to define data quality in your day-to-day work?
6. **Data Domain:** A logical grouping of items, or data, of interest to the organization. Is this the definition you use to define data domain in your day-to-day work?
7. **Data Catalog:** A software application that creates an inventory of an organization's data assets. Is this the definition you use to define data catalog in your day-to-day work?
8. **Data Glossary** (also known as **Business Glossary**): A collection of data related terms described in clear language that everyone in an organization can understand. Is this the definition you use to define data glossary in your day-to-day work?
9. **Data Dictionary:** A collection of names, definitions, and attributes about data elements that are being used or captured in a database or information system. Is this the definition you use to define data dictionary in your day-to-day work?

10. **Data Interoperability:** The ways in which data is formatted that allow diverse datasets to be merged or aggregated in meaningful ways. Is this the definition you use to define data interoperability in your day-to-day work?

11. **Data Asset:** Any entity that is comprised of data. It may be a system or application output file, database, document, or web page. It also includes a service that may be provided to access data from an application. Is this the definition you use to define data asset in your day-to-day work?

12. **Data Science:** Combines math and statistics, specialized programming, advanced analytics, artificial intelligence (AI), and machine learning with specific subject matter expertise to uncover actionable insights hidden in an organization's data. These insights can be used to guide decision making and strategic planning. Is this the definition you use to define data science in your day-to-day work?

13. **Data Owner:** Individual in charge of the data in a certain data domain. A data owner must guarantee that the information inside that domain is correctly maintained across various platforms and business processes. Is this the definition you use to define data owner in your day-to-day work?

14. **Data Product:** An application or tool that uses data to help businesses improve their decisions and processes. Is this the definition you use to define data product in your day-to-day work?

15. **Data Storage and Operations (DataOps):** A collaborative data management practice focused on improving the communication, integration and automation of data flows between data managers and data consumers across an organization. Goal is to deliver value faster by creating predictable delivery and change management of data, data models and related artifacts. DataOps uses technology to automate the design, deployment and management of data delivery with appropriate levels of governance, and it uses metadata to improve the usability and value of data in a dynamic environment. Is this the definition you use to define DataOps in your day-to day work?

16. **Data Archival:** The process of collecting older data and moving it to a protected location so that it can be retrieved if needed in a data forensics investigation. Archives are distinct from backups. With data archiving, the information is moved to free up storage resources. With backups, working data is copied so that it can be restored in the event of a system failure or disaster. Is this the definition you use to define data archival in your day-to day work?

17. **Data Pipeline:** A means of moving data from one place (the source) to a destination (such as a data warehouse). Along the way, data is transformed and optimized, arriving in a state that can be analyzed and used to develop business insights. Is this the definition you use to define data pipeline in your day-to-day work?

18. **Data Lineage:** The process of tracking the flow of data over time, providing a clear understanding of where the data originated, how it has changed, and its ultimate destination within the data pipeline. Is this the definition you use to define data lineage in your day-to-day work?

19. **Data Profiling:** A technology for discovering and investigating data quality issues, such as duplication, lack of consistency, and lack of accuracy and completeness. Is this the definition you use to define data profiling in your day-to-day work?

20. **Metadata Management**: Deliberate, structured data about data. Is this the definition you use to define metadata management in your day-to-day work?
21. **Reference Data**: The data used to classify and reference other data. Is this the definition you use to define reference data in your day-to-day work?
22. **Golden Record Management** (also known as **Master Data Management**): A software/tool that logs/manages golden records. Is this the definition you use to define golden record management in your day-to-day work?
23. **Golden Records** (also known as **Master Data Records**): A single, well-defined version of all the data entities in an organizational ecosystem. Is this the definition you use to define golden records in your day-to-day work?
24. **Extract, Transform and Load (ETL)**: A data integration process that combines data from multiple data sources into a single, consistent data store that is loaded into a data warehouse or other target system. Is this the definition you use to define ETL in your day-to-day work?
25. **Information Assets**: Information relevant to an enterprise's business function, including captured and tacit knowledge of employees, customers or business partners; data and information stored in highly-structured databases; data and information stored in textual form and in less-structured databases such as messages, e-mail, workflow content and spreadsheets; information stored in digital and paper documents; purchased content; and public content from the Internet or other sources. Is this the definition you use to define information assets in your day-to-day work?
26. **Authoritative Data Source**: A repository or system that contains identity information about an individual and is considered to be the primary or most reliable source for this information. In the case that two or more systems have mismatched or conflicting data, the identity information within the authoritative data source is considered to be the most accurate. Is this the definition you use to define authoritative data source in your day-to-day work?
27. **Business Intelligence**: Those systems, tools and technologies used by organizations to manage, visualize and extract valuable insights from their raw business data. Is this the definition you use to define business intelligence in your day-to-day work?

   Selection Options for questions 1-18:
   - o   Yes
   - o   No

Section 2:

Data Stewardship
As a Data Custodian, you need to be able to support data stewards with implementing the technical changes and controls required to support implementation and embedding of **NASA and Federal data policies**.

To help you complete this section, you may need to reference these data term definitions:

- **Information Assets:** Information relevant to an enterprise's business function, including captured and tacit knowledge of employees, customers or business partners; data and information stored in highly-structured databases; data and information stored in textual form and in less-structured databases such as messages, e-mail, workflow content and spreadsheets; information stored in digital and paper documents; purchased content; and public content from the Internet or other sources

28. On a scale of 1-5, how would you rate your ability to complete the following tasks pertaining to data stewardship?

	1: I'm not aware of this task	**2:** I'm aware but need to learn	**3:** I've learned but need coaching	**4:** I can complete this task on my own	**5:** I can teach this task
I am able to write technical requirements relating to how NASA and Federal data policies effect the security and encryption of information assets (Ex. NIST SP 800-53, FISMA, etc)					
I can maintain, manage, understand, and have the ability to document any exceptions to data policies					
I am aware of NASA access controls (Ex. NASA Identify and Access Management system and role-based access controls, MFA, etc)					
I can shape technical requirements for the secure processing, storage, and retrieval of data					

Section 3:

Enhance and Create Data Flows
As a Data Custodian, you need to be able to write and support the execution of technical requirements related to **enhancing and creating data flows** from source to target in application systems landscape to improve data quality.

To help you complete this section, you may need to reference these data term definitions :

- **Data Domain**: A logical grouping of items, or data, of interest to the organization
- **Data Quality**: Data that is accurate, complete, consistent, reliable, up-to-date, and fit for a purpose
- **Data Profiling**: A technology for discovering and investigating data quality issues, such as duplication, lack of consistency, and lack of accuracy and completeness.

29. On a scale of 1-5, how would you rate your ability to complete the following tasks pertaining to enhancing and creating data flows?

	1: I'm not aware of this task	**2:** I'm aware but need to learn	**3:** I've learned but need coaching	**4:** I can complete this task on my own	**5:** I can teach this task
I can conduct system enhancements to embed data quality rules within systems to ensure the data is complete, consistent, accurate, timely, valid and unique for data at rest					
I can conduct data profiling to improve levels of data quality for specific use cases (SQL, Python, R, etc)					
I can conduct statistical analysis and implement statistical techniques such as hypothesis testing, regression analysis, and outlier detection to identify data quality issues					
I can write technical requirements to add data quality rules to data as it moves from one system to another					

Section 4:

Data Integration and Interoperability
As a Data Custodian, you need to be able to coordinate with data stewards and IT to support the technical implementation of **data integration and interoperability**.

To help you complete this section, you may need to reference these data term definitions:

- **Data Domain:** A logical grouping of items, or data, of interest to the organization
- **Data Quality:** Data that is accurate, complete, consistent, reliable, up-to-date, and fit for a purpose
- **Data Integration:** A practice that comprises the practices, architectural techniques and tools for achieving the consistent access and delivery of data across the spectrum of data subject areas and data structure types in the enterprise to meet the data consumption requirements of all applications and business processes
- **Data Interoperability:** The ways in which data is formatted that allow diverse datasets to be merged or aggregated in meaningful ways

30. On a scale of 1-5, how would you rate your ability to complete the following tasks pertaining to data integration and data interoperability?

	1: I'm not aware of this task	**2:** I'm aware but need to learn	**3:** I've learned but need coaching	**4:** I can complete this task on my own	**5:** I can teach this task
I understand how data modeling can be used to improve consistency across systems so data can be integrated more easily					
I am able to identify which common protocols are widely accepted and are well documented (Ex. HTTP and HTTPS) for exchanging data between different systems					
I am able to identify standardized data formats that are easily readable by many different systems and applications (Ex. CSV, JSON, XML)					
I can write technical					

requirements to add data quality rules to data as it moves from one system to another					
I can use data profiling techniques to systematically analyze data to determine data integration requirements					

Section 5:

Reference Data Management
As a Data Custodian, you need to be able to coordinate with data stewards and IT to support the technical implementation of **reference data management**.

To help you complete this section, you may need to reference these data term definitions:

- **Data Domain:** A logical grouping of items, or data, of interest to the organization
- **Reference Data Management:** The data used to classify and reference other data
- **Data Catalog:** A software application that creates an inventory of an organization's data assets
- **Business Glossary:** A collection of data related terms described in clear language that everyone in an organization can understand
- **Data Dictionary:** A collection of names, definitions, and attributes about data elements that are being used or captured in a database or information system

31. On a scale of 1-5, how would you rate your ability to complete the following tasks pertaining to reference data management?

	1: I'm not aware of this task	**2:** I'm aware but need to learn	**3:** I've learned but need coaching	**4:** I can complete this task on my own	**5:** I can teach this task
I can partner with data stewards and data owners to determine requirements for system maintenance for reference data management tooling (Ex. Data Catalogs, Business Glossaries, Data Dictionaries)					
I can support system enhancements related to					

reference data changes					

Section 6:

Metadata Management
As a Data Custodian, you need to be able to coordinate with data stewards and IT to support the technical implementation of **metadata management**.

To help you complete this section, you may need to reference these data term definitions:

- **Data Domain:** A logical grouping of items, or data, of interest to the organization
- **Metadata Management:** Deliberate, structured data about data
- **Data Catalog:** A software application that creates an inventory of an organization's data assets
- **Business Glossary:** A collection of data related terms described in clear language that everyone in an organization can understand
- **Data Dictionary:** A collection of names, definitions, and attributes about data elements that are being used or captured in a database or information system

32. On a scale of 1-5, how would you rate your ability to complete the following tasks pertaining to metadata management?

	1: I'm not aware of this task	**2:** I'm aware but need to learn	**3:** I've learned but need coaching	**4:** I can complete this task on my own	**5:** I can teach this task
I can partner with data stewards and data owners to determine requirements for system maintenance for metadata tooling (Ex. Data Catalogs, Business Glossaries, Data Dictionaries)					
I can support system enhancements related to metadata changes					

Section 7:

Golden Record Data Management

As a Data Custodian, you need to be able to coordinate with data stewards and IT to support the technical implementation of **golden record data management.**

To help you complete this section, you may need to reference these data term definitions:

- **Data Domain:** A logical grouping of items, or data, of interest to the organization
- **Golden Records (**also known as **Master Data Records):** A single, well-defined version of all the data entities in an organizational ecosystem
- **Golden Record Data Management (**also known as **Master Data Management):** A software/tool that logs/manages golden records
- **Extract, Transform and Load (ETL):** A data integration process that combines data from multiple data sources into a single, consistent data store that is loaded into a data warehouse or other target system
- **Data Lineage/Data Flows:** The process of tracking the flow of data over time, providing a clear understanding of where the data originated, how it has changed, and its ultimate destination within the data pipeline

33. On a scale of 1-5, how would you rate your ability to complete the following tasks pertaining to golden record data management?

	1: I'm not aware of this task	**2:** I'm aware but need to learn	**3:** I've learned but need coaching	**4:** I can complete this task on my own	**5:** I can teach this task
I am able to identify authoritative sources for Extracting, Transforming and Loading (ETL) in order to create a golden record					
I can conduct an impact analysis and understand data lineage (data flows) for my respective NASA Office, and the impact of technical changes					

Section 8:

Additional Feedback

Your feedback matters - please let us know if there are additional skills for Data Stewards we should include, any concerns or questions you would like to share with the team, or any other general feedback.

34. Please provide any additional feedback below:

Answer: _____